Large Scale Observations: a SEARCH workshop
November 27-29, 2001, Seattle, WA
http://www.epic.noaa.gov/SEARCH/obs/workshop/

Reanalysis; The Observations and Analyses

(Roy Jenne, NCAR, Oct 2001)

The subject of reanalysis includes most of the world's weather observations of surface and upper air conditions from 1948 – 2001. Also, this subject includes the models used for reanalysis, the reanalysis projects, the output from the projects, and where to obtain the data. We can only give an overview of this large subject in this paper, and point to some of the sources for more information.

First we will outline some of the world's early work on preparing datasets of observations. The work at NCAR from 1991 – 2001 to prepare the observations for reanalysis will be summarized. Then the reanalysis methods and the reanalysis projects will be briefly covered. There is more data for the pre-1948 period than people realize, but much has not been digitized and some has been lost. This subject is briefly touched on, because there is a strong interest in learning more about climate oscillations back to 1900 or earlier.

DATA PROJECTS AND ARCHIVES THAT LEAD UP TO REANALYSIS

The Data Support section at NCAR was established in 1965. At first we started gathering hemispheric analyses from weather centers such as NCEP and the Navy. Later, we started gathering datasets of GTS observations from NCEP and the Navy. There also had been huge projects by the USAF in the 1950s and 1960s to gather worldwide observations and to key-enter it. They produced several hundred million data cards. Also, NCDC key-entered millions of punched cards. Asheville probably had about 600 million punched cards by the end of the 1960s. The cards were put onto tapes in the 1970s and then the data became a little more affordable. It would have helped us at NCAR to be able to obtain original versions (on tape) of more of the data.

At NCAR we had also gathered observations from several countries that had national and foreign data holdings (such as Britain and France). In 1990, our NCAR Data Support Section made an agreement with the model development group at NCEP (Eugenia Kalnay) to team together to produce a long reanalysis. By then we already had a lot of the necessary observations, but more processing was needed.

WORK BY NCAR, 1991 – 2001, TO HELP REANALYSIS PROJECTS

NCAR needed to prepare the observations so that they could be more easily used for reanalysis. We ran many diagnostic checks to identify and correct systematic errors. Examples of these are (1) stratospheric data always assigned to the wrong pressure levels for several years, (2) the date/time was offset by one day or more for 100 station-years, and (3) locations of aircraft recon wrong by 10 degrees of longitude, etc. NCAR did a lot of work to develop better location and elevation data for all of the upper air stations. The better locations were attached to every report, by using a format

shell for reanalysis. We did hydrostatic checks on all of the rawinsonde data in order to locate systematic errors, but we only changed part of the random errors (some of the more obvious ones). It is better to leave the control of random errors to later stages of the data processing. NCAR gathered observations directly from more countries during 1991 – 2001. We also obtained some more of the old, big component datasets from USAF and NCDC units at Asheville.

NCAR sent our processed sets of observations to NCEP during 1993 – 2001. NCEP converted the data to Bufr format, and then merged component datasets to eliminate duplicates. They then ran the observations through their latest QC programs, which are always under further development. The component datasets will be archived because merging and dupe-elimination are complicated processes. The same observation often will not be the same in two or more component datasets, and its date-time stamp may differ by one or two hours.

PREPARE AND IMPROVE SEVEN DATASETS OF WORLD OBSERVATIONS

The seven main datasets of world observations were needed in order to actually make new analyses of the world's atmosphere each 6-hours for 1948 – 2000 (Table 1). The data were in many national archives, and in a number of larger subsets. We needed to obtain the data (not easy), and do many, many diagnostic checks to find systematic errors, and remove them. The work started in 1967 at NCAR and became much more intense in 1991 when our work at NCAR DSS started to prepare for the NCEP/NCAR long reanalysis. The datasets of observations are now much better than anything that the world has ever had. There is still data (hard to get) that should be added, especially for earlier years.

This data preparation has been a huge amount of work, and a nice benefit for world science. It is encouraging that it has worked out well. These observations and new reanalysis output will be key resources for new reanalysis projects and for other world research during 2001 – 2010.

TABLE 1. SEVEN MAIN SETS OF WORLD OBSERVATIONS
For each category of data, NCAR obtained and processed many separate data inputs.
These are being used to prepare new global analysis at NCEP, ECMWF and NASA.
The rawinsondes are from land stations, ships at sea, and dropsondes.

	Data Years	Number Years	Start Work At NCAR	Recent Work	Comments
Rawinsondes	1946-on	55	1967	1991-01	Some earlier data
Pibals	1942-on	59	1973	1991-01	Some earlier data
Aircraft	1947-on	54	1973	1992-00	None in 1959
Sat cloud winds	1967-on	34	1973	1991-95	More 1973-on
Satl soundings	1969-on	31	1973	1991-00	Better 11/1973-on
Sfc 3-hr synop	1946-on	55	1976	1992-01	More obs 1967-on
COADS ocean sfc	1854-on	145	1981	1988-01	Some earlier data

THESE 7 DATASETS ARE THE MAIN SUCCESS STORY FOR DSS.

- They describe the world weather history for 53 years.
- And they make it possible to do reanalysis projects.
- And much other research uses observations.

THE COVERAGE OF OBSERVATIONS

NCAR has coverage plots of observations used in reanalysis. For upper air data, we also have long tables that show the number of rawinsonde reports, and separately, the number of wind-only (pibal) reports received each month-year from each fixed location. Sometimes the communications systems did not capture all of the observations that were taken (as from US rawinsonde stations in Antarctica). It is useful to browse these tables to see if there are problems, and to think of methods to increase the data counts if they are low for isolation stations.

It is true that the number of reporting stations increases from 1948 – 1975. However, there are cases where the earlier observing networks are better. There was much better coverage of permanent ship rawinsondes for oceans in the northern Hemisphere during 1948 – 1973 then in later years. Also, there were many ocean area recon flights during 1947 – 58. In future reanalyses, we should test to see whether the isolated data is used in the best way, and all of the old dropsondes should be digitized.

PROGRESS IN DEVELOPING THE REANALYSIS DATASETS OF OBSERVATIONS

The NCEP/NCAR reanalysis used Version 1 of all of the observations. NCAR sent the last Version 1 of all of the 1948 – 57 block of years to NCEP by 03/1998. At that time, Version 1 was done. Version 1 was a very good dataset, but NCAR could still add more data and improve the knowledge about the location of stations. Version 2 was delivered to ECMWF by about 02/1999, but NCAR was still working on many new data projects at that time. Most of the data for Version 3 was sent to NCEP by June 2001, and sent again in July 2001 (because of mass store problems at NCEP). The ECMWF ERA-40 reanalysis will be mostly based on Version 3 data. The NCEP-2 reanalysis (1971 – 2000) was done with a slightly improved set of Version 1 data.

THE POWER OF THE REANALYSIS METHODS

The reanalysis procedures make use of all of the relevant types of data. Rawinsondes, aircraft, satellite data, and surface data all help. We need enough data (with fairly good error characteristics) to obtain good analyses. The forecast models have also been improving a lot in the past twenty years. If the analyses are very good, then the forecasts will be good. The forecast (first guess) are the main basis for the analyses where there are no new observations. Therefore, a good coverage of observations over Asia will tend to give good forecasts over the north Pacific, and therefore rather good analyses there, too.

When observations become too sparse, and too far away from good data coverage, then the quality of the analysis will go down. In this case, a better analysis could be produced by using 4D-var methods. The present 3D-var methods only use forward time continuity. But if an unknown storm hits an isolated island, it must have also existed in the past. The 4D-var methods build both forward

and backward continuity (for a day or so) into the analysis procedure. We expect that some 4D-var methods will be used in the future, especially for early years.

THE MAIN REANALYSIS PROJECTS

In the planning for the NCEP/NCAR reanalysis, we first said that we would do the years 1958 – on. It was possible to gather a number of observations from 1948 – on. During about 1996, the panel extended the plans to include the years 1948 – 57; this was a good decision. We note that during the early years, the standard time for the world's rawinsonde and pibal data was 03 and 15 GMT; it became 00 and 12 GMT starting 06/1957. These time changes had to be accounted for in the reanalysis procedures. Table 2 gives information about the main reanalysis projects. The reanalysis output can be used to help make seasonal forecasts. That is a motivation to keep updating the reanalysis each month as is done by NCEP. The production of the reanalysis at NCEP was first done on a Cray YMP computer with 8 processors. Then it was changed to a Cray J90 with 16 processors. Each of these computers could deliver a steady 1.0 Gflop of real computer power and that was enough to accomplish about 20 to 30 days of 6-hourly global analyses during each day, at the resolution of T62, L28. The resolution of the reanalysis is given in the table below. T62, L28 means 62 spectral waves and 28 levels. The horizontal resolution in kilometers is also given in Table 2.

TABLE 2. MAIN REANALYSIS PROJECTS

A few people helped to develop the idea of reanalysis projects during 1986 – 90. The development work for two early projects was done during 1991 – 94 and production started 06/1994. The projects are listed in the table below. The resolution of the reanalysis is given in the table below. T62, L28 means 62 spectral waves and 28 levels. The horizontal resolution in kilometers is also given.

Project	Model	Years	Km	Data Dates	Main Production	Updates
1. NCEP/NCAR	T62, L28	54	208	1948 – 2001	06/94-07/98	Yes
2. NCEP-2	T62, L28	23	208	1979 – 2001	05/98-12/00	Yes
3. ECMWF (ERA-15)	T106, L31	15	125	1979 – 1993	06/94-09/96	No
4. ECMWF (ERA-40)	T159, L60	47	83	1957 – 2003	07/00- ?	
5. NASA, Goddard	2.5x2, L20	17	240	1979 – 1995	~1994-98	No
6. Japan	e T106, L30	26	125	1979 – 2004	?	?

THE USE OF SATELLITE SOUNDER DATA IN REANALYSIS

The two main series of sounder data are from VTPR (11/1972 – 02/1979) and TOVS (11/1978 – 2001). The VTPR has 8 IR channels; TOVS has many IR channels, VIS, microwave (MSU), and separate stratospheric radiances (SSU). The VTPR radiances should be good from 11/1972 – on, but the calculated soundings are bad until 03/1975. The NCEP/NCAR 50-year reanalysis used sounder data from 03/1975 – on. NCEP ran reanalysis experiments during the VTPR and TOVS overlap period. They made analyses using (1) VTPR, (2) TOVS, (3) no sat. The RMS differences between 500 maps using VTPR or TOVS were small. The daily no-sat analyses had rather large differences from the others.

In the ECMWF ERA-40 reanalysis, the radiances are being used directly (not the pre-existing soundings). They will use the full record of data from VTPR and TOVS, thus 11/1972 – on. The operational analyses at both NCEP and ECMWF have been directly using radiances for several years.

The world's first satellite sounder (SIRS) started in 04/1969. I hope that people will also try to use some of the earliest data in the future, but that task would be more difficult because of resolution, and other factors.

A PROBLEM IN TOVS USE HAS BEEN FIXED (1997 – 2000)

NCEP has been producing the NCEP/NCAR reanalysis (1948 – 2001) and the NCEP-2 reanalysis (1979 – 2001). They update both reanalyses each month. NCEP compares the two analyses with each other. They found that they compared closely for many years, but in Mar 1997, they diverged. NCEP found that in the NCEP/NCAR reanalysis they had started to use TOVS soundings over land areas by mistake. NCEP has rerun all of the NCEP/NCAR reanalyses from 1997 – on to exclude TOVS over land, as done before. This task was completed in 2001. Then NCEP reran the other products such as 8-day forecasts done each 5th day. New annual CD-ROMs for 1997 – 2000 will also be produced.

THE USE OF TOVS DATA IN NCEP REANALYSES

The two reanalyses (NCEP/NCAR and NCEP-2) use TOVS sounder data that has been calculated before the reanalysis is run. This means that the cloud clearing of the radiances, and the calculation of the vertical temperature structure is done external to the analysis process.

NCEP uses TOVS sounding data over the whole world in the stratosphere. Over land areas they exclude TOVS sounding data under 100 mb. When there is enough rawinsonde data, the TOVS data does not help; it is a negative factor then. Over the Arctic, NCEP uses all of the TOVS sounder data in reanalysis.

THE OUTPUT FROM THE NCEP/NCAR 53-YEAR REANALYSIS

The output from the NCEP/NCAR reanalysis (now 53 years) and the NCEP-2 project (now 1979 – 2000, 22 years) will be described. The data outputs are given each 6-hours for many years. The analysis and forecast models use a smooth elevation of the earth. The analysis is done in sigma coordinates, at model resolution (208 km). The sigma coordinates follow the undulating surface terrain. Consider the four lowest model levels above the ground. These are at mid-level sigmas of 995.00, 982.08, 964.37 and 942.55. These sigmas correspond to the following approximate distances above the ground: 41 m, 147 m, 292m and 470m.

The models keep track of radiation, clouds, precipitation, etc. These terms are in the flux fields. Also, winds at 10 m are calculated, plus temperature at 2 m. The atmospheric data are interpolated from the sigma coordinates to the more familiar pressure coordinates. Please see Kalnay, et al., 1996, for tables of all of the outputs. A popular portion of the data is on one CD-ROM for each year. And monthly statistics are on one CD-ROM for many years. NCAR has been delivering huge amounts of reanalysis output data to users.

TABLE 3. OUTPUTS FROM THE NCEP/NCAR REANALYSIS

		Volume/Year
1.	Pressure level data (17 levels), 2.5° resolution	2.53 GB
2.	Flux fields: surface temp, clouds, radiation, etc., 208 km resol.	1.99 GB
3.	Data on sigma levels, 208 km resol.	4.91 GB
4.	Monthly statistics (part 208 km, part 2.5°)	0.27 GB
5.	All outputs	54. GB

EARLY DATA THAT MAY HELP ARCTIC STUDIES

Many national surface weather-observing networks had rather good coverage of synoptic observations by at least 1880. The USA had rather good coverage of pressure data from Denver to the East by 1850. The NCAR library has daily, published Northern Hemisphere weather maps for 1899 – 1939 and then other documents continue the map series. The observations and pressure contours are plotted on the maps. The NCAR Data Support Section has grid points from the daily maps for 1899 – on. These are usually on a diamond grid, 5° Lat by 10° Lon. There is a long list of data sources at the front of each monthly publication (I am looking at Dec 1903 now). NCAR has the COADS dataset of ocean ship and buoy observations for 1854 – on. It would help to make more of the early land observations available in digital form. One possible project could be to use the existing daily SLP grids as a guess, and then also use available observations for land and ocean areas. Also, the existing daily maps could be digitized at a better resolution. The SLP grids and station pressure and temperature data could be used to make estimates of airflow at 700 mb. Then the existing pibal winds and early raobs could be used to help correct those estimates.

GUIDE TO DOCUMENTS ABOUT OBSERVATIONS, ANALYSES, ETC. AT NCAR

During the reanalysis projects, NCAR prepared many documents to describe the availability of data-sets, and to document individual collections of data. Also, we needed to inform NCEP and ECMWF about the observations that we were sending. In Apr 1999 we started a document project to gather the smaller papers into larger documents and to write new text. The production scanning started 03/2000. By Oct 2001 we had scanned 9589 pages. Table 4 lists some of the categories of documents.

GUIDE TO DATA AND DOCUMENTS ON THE WEB

Go to the NCAR Data Support home page and look for "Reanalysis: Guide to Global Observations and Output." (http://dss.ucar.edu)

TABLE 4. GUIDE TO SCANNED DATA DOCUMENTS AT NCAR

Documents have been prepared, gathered, and scanned in our Data Support group at NCAR.

		Paper Bundles	Pages
1.	Papers about observations for reanalysis	6	719
2.	Papers about rawinsonde and wind only data Includes ~634 pages of manuals	16	e1330
3.	Documents about surface land observations	8	e260
4.	Papers about satellite data	10	~600
5.	Papers about aircraft data and balloon data		
6.	Papers about data from projects		
7.	Energy, climate and population; the connections	9	658
8.	Total scanned docs in Oct 2001	~145	9589

MAIN PAPERS ABOUT THE NCEP/NCAR 50-YEAR REANALYSIS

Kalnay, et al, 1996, have a long paper about the NCEP/NCAR reanalysis in the Feb 1996 issue of *Bull. AMS*. This describes the projects and includes many tables that list the output variables.

The first WCRP conference on reanalysis had many papers about the reanalysis output, printed in WCRP-104 (conference held in Oct 1997). The second international conference on reanalysis was held in Aug 1999, and the papers are printed in WCRP-109.
References:

Kalnay and co-authors, 1996: The NCEP/NCAR 40-Year Reanalysis Project, *Bull. AMS, Vol 77*, pp 437 – 471.

Kistler and co-authors, 2001: The NCEP-NCAR 50-Year Reanalysis; Monthly Means CD-ROM and documentation. *Bull. AMS, Vol 82, No. 2,* pp 247 – 268 (Feb 2001 issue).

US Joint Met Com, ~1943, *Historical Weather Maps, Northern Hemisphere Sea Level* daily maps. Monthly publ, 1899 – 1939

WCRP-104, 1998, First WCRP International Conference on Reanalysis, held in MD, USA, Oct 1997. Proceedings from WMO, 480 p.

WCRP-109, 2000, Second WCRP International Conference on Reanalysis, held in UK, Aug 1999. Proceedings from WMO, 465 p.

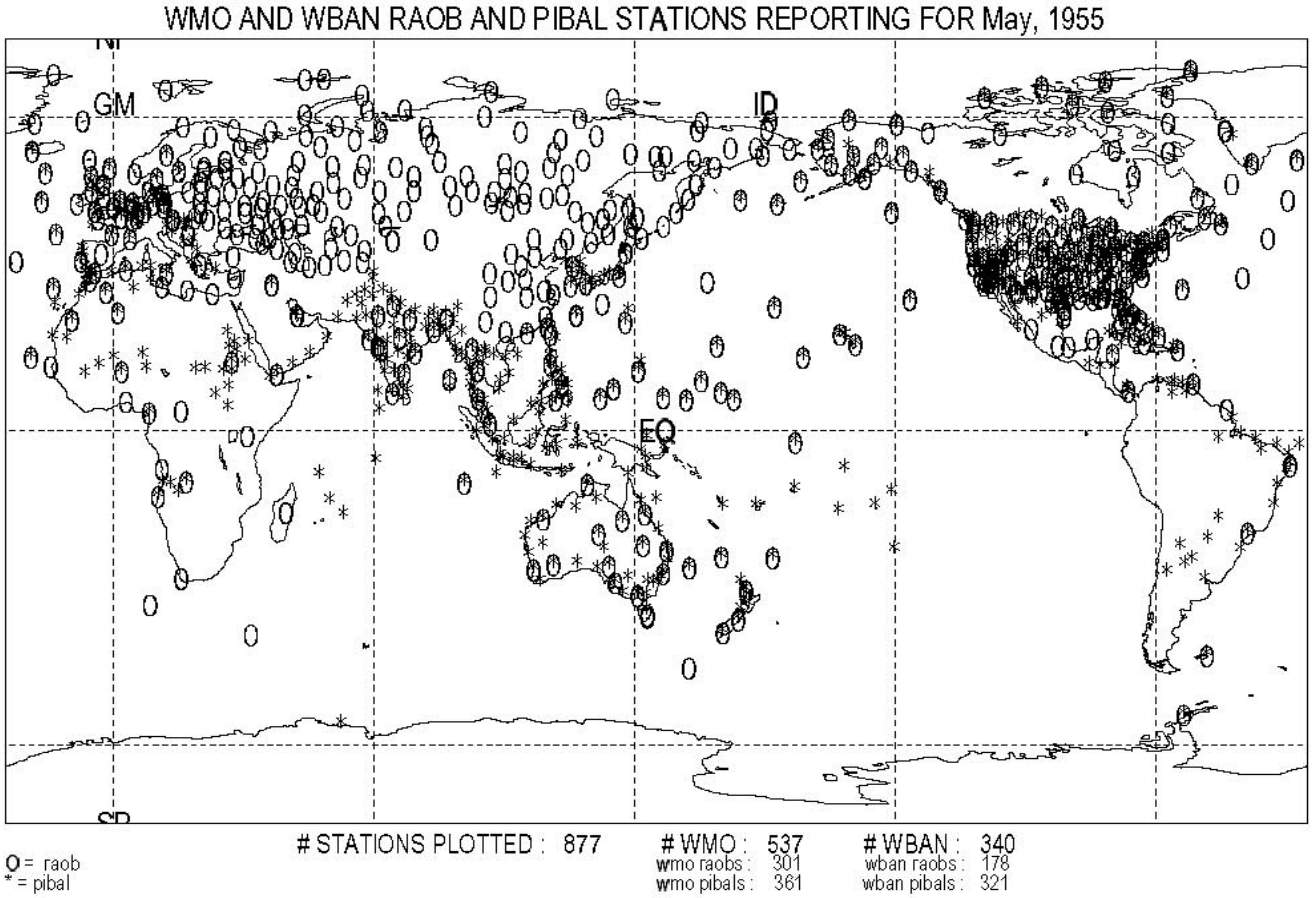

FIGURE 1. Coverage plot of world rawinsonde and wind-only upper air data for May 1955. The 0's are rawinsondes and are pibal stations. Data for China starts 01/1954. The data coverage for Canada becomes excellent starting 07/1955.